

米国のエビデンス仲介機関の機能と課題

－米国 WWC 情報センター (What Works Clearinghouse) の例より－

Roles and Issues of the Agency Working with Evidence

－the example of What Works Clearinghouse in the United States－

豊 浩子*

YUTAKA Koko

Abstract

What Works Clearinghouse (WWC) is a website established by Institute of Education Sciences at the U.S. Department of Education under the current American educational reform, endorsed by the No Child Left Behind Act. With its mission to provide educators, policy makers, researchers, and the public with a source of scientific evidence on effective education programs and practices, WWC has been working to review almost all available research on the interventions in various topic areas in education. WWC has a rigorous protocol and procedures for the assessment of evidence in each topic area, and normally admits only randomized control trials. The current topic areas reviewed by WWC are: Adolescent Literacy, Beginning Reading, Character Education, Dropout Prevention, Early Childhood Education, Elementary School Math, English Language Learners, Middle School Math, Students with Learning Disabilities, High School Math, and Early Childhood Education for Children with Disabilities. The publications produced by WWC include Topic Report, Practice Guide, Intervention Report, and Quick Review. Although WWC has produced many publications to disseminate its review and assessment on educational research, there are some critiques that WWC's focus on randomized control trials has raised some serious issues about the meta-analyses conducted by WWC. The current educational reform originally aimed to equip all American students with a strong academic achievement including the critical thinking, problem solving skills that the globalized economy needs. However, the accountability system, which the reform demands of schools and districts, seems to have let them search for "quick answers" that help them teach students just the basic skills required by the high-stakes standardized tests. Thus, the current situation may take the reform in unpredicted directions, and it is uncertain that WWC will continue to play the same role in the near future.

1. はじめに

近年、欧米の先進諸国では政府が主導して、研究と政策をつなぐエビデンスの収集・レビュー・分析・評価を行う仲介機関が設立、運営される例が見られる (OECD 教育研究革新センター 2007)。本論では、米国の What Works Clearinghouse (WWC 情報センター, 略称 WWC) のケースをもとに、エビデンスの仲介機関の機能とその課題について述べることにする。

* 研究協力者

WWC は、米国教育省下の Institute of Education Sciences（教育科学研究所、略称 IES）によって設立された、ウェブサイトによる、教育の評価研究に関するレビューのデータベースである（<http://IES.ed.gov/ncee/WWC/>）。

その目的は、教育における介入の効果について、科学的エビデンスとして信頼性の高い情報源を提供することであり、その活動の中心は、介入の効果にかかわるエビデンス、特に因果関係が認められるエビデンスのレビュー及び要約を行うこととされ、その役割は、WWC が対象とするトピック分野におけるあらゆる研究報告を網羅し、厳密な基準によってそれらのエビデンスの信頼性を判断し、取捨選択することにあるとされる。（Boruch & Herman 2007）。

Boruch & Herman（2007）によれば、WWC が対象としている利用者は、政策立案者、教員ら教育関係者、保護者、研究者、一般市民に及ぶ幅広い層である。

2. WWC 設立の背景

(1) 米国の教育改革の動向

米国では 1980 年代、グローバル化による国際経済競争の激化及び米国の国際経済的地位の低下に対する危惧と不安を背景に、『A Nation at Risk（危機に立つ国家）』が発表され、「Back to basics（基礎に帰れ）」の教育改革に向かうことになった（荻谷 2001, Shoenfeld 2004）。

この動きは、従来、各州レベルで行われてきた教育政策に対し、米国の将来がかかっている教育の内容・質を米国全体で向上させる必要があるという認識のもと、連邦政府が主導して教育改革を行う流れへとつながった。

1989 年、当時のブッシュ大統領は州知事や財界のリーダーを招いた第 1 回「教育サミット」を開催し、そこで「全米教育目標」の達成戦略「2000 年のアメリカ」が設定され、各州は教育スタンダードの設定と学力テストの実施を求められた（岸本 1998, 杉浦 2006, 松尾 2010）。

この、全米挙げての取り組みは、連邦政府の呼びかけによる、各州でのスタンダード開発とそれに基づく教育改革へと向かった。

続くクリントン大統領もこの流れを踏襲し、1994 年に教育改革に関する初の連邦法 Goals 2000: Educate America Act（2000 年の目標・アメリカ教育法）を成立させ、Elementary and Secondary Education Act（初等中等教育法、略称 ESEA）を修正して Improving America's School Act（アメリカ学校改善法）を制定した。（杉浦 2006）

以上のように、政権を超えた教育改革への取り組みが続いたにもかかわらず、1990 年代の終わりになっても、「全米教育目標」の達成や、生徒の学力の向上といっためざましい改善は見られなかったとされる（松尾 2010）。

2001 年、ブッシュ政権で、クリントン政権の教育政策が引き継がれる一方で、ESEA を改正した the No Child Left Behind Act（落ちこぼれをつくらないための初等中等教育法、略称 NCLB 法）が制定され、2002 年に施行された。

NCLB 法は、米国における学力格差の是正を図ると同時に、科学的に実証された質の高い研究、すなわち科学的エビデンスを、教育における政策決定に利用することを重視するとされた（田辺 2006）。

2002 年には、The Education Sciences Reform Act of 2002（教育科学改革法）によって、教育省下に

新たに IES が設立された。IES はそのミッションを、教育実践と政策が基盤とする厳密な証拠を提供することとし、何がなぜ有効か、有効でないか、を明らかにし、全ての生徒、特に、教育的リスクの高い生徒のための教育成果の向上を目指すとしている¹⁾。

IES の活動としては、学力達成度向上の方法に関する調査研究の助成、国の教育プログラムに関する大規模評価、National Assessment of Educational Progress (全国教育プログラム評価、略称 NAEP) のような教育状況に関する様々な統計報告などが挙げられている。IES はまた、エビデンスを基盤とする教育のためのインフラ構築のサポートを行うために、州に、長期データのシステム開発の助成金を出している。さらに、WWC や、ERIC 教育データベース、会議や出版物、10 の地域教育ラボラトリー、10 の国立研究開発センターを通して、一般市民や教育実践者に対する、教育に関するエビデンスの普及に努め、法的及び文化的に偏った政治的影響は受けないとしている。

WWC は、このような動向の中で、2002 年、IES によって設立された。

Boruch & Herman (2007) によれば、WWC は、IES とキャンベル共同計画と American Institute for Research とのジョイントベンチャーに対する契約を通じて資金援助され、その目的と運営方針は契約により明確にされた。具体的には、IES の入札に 3 社のリサーチ会社が応じ、American Institute for Research が契約を得、2440 万ドルの資金援助のもと、活動を行ったとされている²⁾。

2007 年から 2012 年までの 5 年間は、Mathematica Policy Research Inc. が 5000 万ドルで契約を結び、調査を委託されて行っている (Slavin 2008)。この Mathematica Policy Research Inc. は、ランダム化比較試験 (Randomized Controlled Trials, 略称 RCT) 実施のノウハウと実績を持ち、当社のウェブサイトによれば、特に障害、幼児期、教育、雇用、医療、栄養に関する政策について専門性が高いコンサルティング企業、となっている³⁾。

WWC は契約先の変更及び資金の増額に合わせて、公表されるレポートの迅速化が図られ、また、より実践的な内容のレポートも発表されることになったようである⁴⁾。2008 年からは、全国の一般メディア上で報道された最新の研究について迅速にレビューを行う「クイックレビュー」(Quick Review) を公表するようになった⁵⁾。また、教員や、学区、州の教育関係者らのために、实际的で使いやすい内容を目指した「実践ガイド」(Practice Guide) も発表されるようになった。

Boruch & Herman (2007) が、WWC のウェブサイトは透明性の高い最新の基準を掲げ、進化し続けている、と述べているように、WWC の刊行物とそのウェブサイトは、できるだけ利用者に使いやすく参照しやすい内容と形式に修正されてきているようである。これは、エビデンスの仲介機関として、次に述べる、科学的エビデンスの普及を推進する連邦教育省の方針に沿ったものだともいえよう。

(2) 教育政策における科学的エビデンスの推進

連邦教育省は、2001 年の NCLB 法の制定以降、教育研究における「厳密なエビデンス」に裏付けられた、科学的に根拠のある方法論の活用を推進してきているが、その背景には、教育における介入の有効性について、信頼におけるエビデンスに基づいて因果関係が明示されるべきである、という考え方があるとされる (Boruch & Herman 2007)。

WWC はエビデンスの仲介機関として、教育省の要求する「厳密なエビデンス」を教育分野における大量の研究を網羅して探索、レビュー、評価を行い、その結果を利用者にわかりやすい形で報告する、という役割を担っている。

教育省はまた、一貫して「厳密なエビデンス」とは RCT によるものだと定義している。2003 年には、科学的エビデンス、また scientifically-based research（「科学的根拠のある」研究）の定義に関して、「厳密なエビデンスによって裏付けられた教育実践の識別と実施：ユーザーフレンドリーガイド」というガイドラインを発表し、「科学的根拠のある」エビデンスとは RCT によって得られたエビデンスのことを指すと明言し（田辺 2006, Boruch & Herman 2007）、以降もこの方針の推進のために、ガイドラインや出版物を通し⁶⁾、研究の計画・実施や教育施策への取り入れにも RCT の採用を推奨している。

これは、RCT であれば、調査の初めの段階では、子どもであれ、家族であれ、学校であれ、介入にかかわる対象は、より効果的とみなされる他の介入にかかわる対象と格別異ならないため、RCT からは公正な比較が示されるという見解に基づいているとされる（Boruch & Herman 2007, Coalition for Evidence-Based Policy 2003）。

RCT は、医学分野では 1950 年代から、そして 1970 年代からは雇用、研修、福祉分野の研究においても、優位を占めるようになったが（田辺 2006, Boruch & Herman 2007）、教育の分野ではそれほど重視されてこなかった。これには、RCT の実施が教育の現場では容易ではなく、また実験の対象となるグループとその比較グループを決定することが倫理的にも問題があるなどの理由から、積極的に推進されてこなかった背景がある。教育省の方針はこういった従来の傾向を大きく変更するものであった（田辺 2006, Boruch & Herman 2007）⁷⁾。

教育省の「科学的根拠のある研究」を推進する姿勢は、同省の 2002-2007 年及び 2007-2012 年の長期計画目標でも明確にされている（U.S. Department of Education 2002, 2007）。

WWC 及び IES には RCT を強制する力はないとされてはいるが（田辺 2006）、実際のところ、教育省の 2002-2007 年の長期計画目標では、科学的根拠に基づく教育研究への助成方針が打ち出され、研究助成金の申請書やガイダンスは、研究知見を反映するように改訂され、研究助成金受給者は、有効性が実証された方法論を用いるよう義務づけられた。（U.S. Department of Education 2002）。こういった動きを反映し、教育の分野での RCT による調査研究が増加しているといわれる（Slavin 2006）。

教育省は、2007-2012 年の長期計画目標では、こういった努力を引き続き行うことを明言し、教育をエビデンスに基づく分野へと変える、という目的を掲げ、下記の内容を宣言している。

- (1) 読み書きの達成、数学と理科の達成、教員の質と有効性を向上させるための、効果的なプログラムと実践を開発または認識する。

IES はそのために、the National Center of Education Research（国立教育研究センター）及び the National Center of Special Education Research（国立特殊教育研究センター）を通じて、質の高い研究プロジェクトに対する助成金の支給を続ける。研究プロジェクトが効果的なプログラムや実践を生み出しているかどうかの判断については、WWC の厳密なエビデンス基準が適用される。

- (2) 教育プログラムと実践の有効性についての情報を普及する：IES は、教育における意思決定者に対する、教育プログラムと実践の有効性に関するリソース提供のために、WWC に対する助成を続ける。

IES はまた、WWC による「実践ガイド」の開発にも助成を行う。WWC の「介入レポート」が個々のプログラムや実践に焦点を絞っているのに対し、「実践ガイド」は、カリキュラム、教員研修、評価、アカウンタビリティなどの分野にわたる、広範囲の解決を要する特定の問題に関する

実践について、首尾一貫したガイダンスと提案を提供する。「実践ガイド」は入手可能なベストのエビデンスを取り入れ、特定の提案を支持するエビデンスの質について利用者に注意を促す。「実践ガイド」は連邦教育省によって広く普及されることになっている。

(以上、U.S. Department of Education 2007 より引用)

上記のように、研究プロジェクトに対して、助成金を支給するか否かを決定する評価には、WWC のエビデンス基準が適用されることになっている。

また、WWC と、WWC の発表する「実践ガイド」に対して、IES から助成と支援を行うことも明確に打ち出され、エビデンスの仲介機関としての WWC の役割に対する期待が引き続き大きいことが示唆される。

なお、この 2007-2012 年の長期計画目標では、WWC で有効性を確認されて教育省から助成を受けるプログラムや実践の数を増やすと同時に、WWC のウェブサイトの閲覧件数を 2006 年の 680 万回から、2012 年には 1500 万回にしたいとして、表 1 のように目標に挙げている (U.S. Education Department 2007)。

表 1：連邦教育省によって発表された、WWC 基準で有効とのエビデンスを証明されて教育省の助成金を受けているプログラムや実践の数と WWC の利用についての目標値

	基準	目標					
	2006年	2007年	2008年	2009年	2010年	2011年	2012年
WWC 基準で有効とのエビデンスを証明されて教育省の助成金を受けているプログラムや実践の数							
・読み書き	3	6	11	13	15	17	20
・算数または理科	1	3	7	10	12	15	18
・教員の質	1	3	5	7	10	12	15
WWC ウェブサイトの閲覧件数	680万	800万	920万	1050万	1200万	1350万	1500万

(U.S. Education Department 2007)

上記から、教育の分野において、WWC 基準によるエビデンスを証明されて教育省から助成金を受けるプログラムや実践の数によって、エビデンスにおける WWC 基準がどの程度、教育の実践の場に浸透しているか、をある程度見ることが可能と考えられている様子が窺える。

また、WWC のウェブサイトの閲覧件数が、WWC が実際にどの程度利用されているかを測る指標として重要視されていることから、教育省が、WWC の利用をできるだけ広範囲に普及させ、教育研究エビデンスの仲介機関としての活用拡大を推進する方針であることがわかる。

初期の WWC に対しては、実施するレビューがわかりにくい、レビューのスピードが遅いといった批判も出たようであり⁸⁾、その対応策の一環としてか、教育省は、「クイックレビュー」や「実践ガイド」といった、より迅速でわかりやすい刊行物を追加する他に、Doing What Works という、別のウェブサイトも設立している (<http://www.dww.ed.gov/site/>)。このサイトは、WWC のレビュー内

容を基に、教員や、学校、学区や州の教育関係者に対して、よりわかりやすい情報を提供することを目的としているようであり、特に「マルチメディア概観 (Multimedia Overview)」として、音声と映像によって内容を提供している。Doing What Works のサイトは WWC のサイトでも、パートナーサイトとして紹介されている。

以上の動向からも、連邦教育省の、科学的に根拠のあるエビデンスに基づく教育研究を推進する戦略に際して、WWC がエビデンスの仲介機関として確固たる位置を占め、その推進の中核となっている様子が窺える。

(3) 科学的根拠の基準としての RCT

WWC は、教育の場におけるプログラム等の介入に関して、教育省の方針を反映し、何が有効か (What works) を判断するための質の高いエビデンスとして、特に RCT によるエビデンスを重要視する方針を強調している。WWC の運営上、RCT は、因果関係を提示する上で準実験より信頼性が高いとして最優先される一方、準実験は優先順位が低く、WWC によるどの報告においても、科学的エビデンスの基準がより低いものとして扱われている。

RCT を最優先するという連邦教育省の決定は、教育や評価の分野で論争を巻き起こすことになった (田辺 2006, 佐々木 2010, Donaldson & Christine 2004)。

評価研究の第一人者であるとされるスクリヴェン (Michael Scriven) は、RCT が因果関係を証明するための最良のデザインであるという主張を、一貫して鋭く批判してきたとされる (佐々木 2010)。そして、連邦教育省の IES が、NCLB 法において RCT をほとんど唯一の望ましい調査デザインとする方針を全面的に打ち出したことに対して、スクリヴェンが教育省の方針に対する反対の急先鋒となり、アメリカ評価学会の意見表明として、RCT のみが有効なエビデンスとして扱われることに対する反対意見をとりまとめて提出した。この動きに対し、同評価学会内で、教育省の立場を支持する人々が別の意見表明を提出し、スクリヴェンらに対して真っ向から反対する動きとなり、アメリカ評価学会が二分されたとされる⁹⁾ (佐々木 2010, Donaldson & Christie 2004)。

アメリカ評価学会ではその後、双方を招いた公開討論が行われたが、スクリヴェンはその際、RCT が優れているという考え方を否定しているのではなく、教育省が、他の調査デザインを排除して RCT のみを優先していることに対して反対しているのだ、と述べている (Donaldson & Christie 2004)。このように、教育省の方針に反対する立場であっても、RCT の優位性を否定しているのではなく、RCT が唯一のエビデンスとして扱われることに異議を唱えているのである (田辺 2006, Slavin 2008)。

教育省を支持する立場の代表として意見を述べたリップシー (Mark Lipsey) は、RCT 以外の調査方法を認めないといっているわけではない、ただ、それ以上に介入評価に説得力を持つ調査方法が出てこないだけだ、と述べている。そして、保健や医療その他、教育以外の分野では RCT の有効性が高く評価、理解され、広く利用されており、今更議論にもならないが、教育関係者の間ではイデオロギーとして RCT が受け入れられにくいのだと述べた。

スクリヴェンもリップシーも、RCT がプログラムの因果関係を評価する現在可能なベストの方法論であり、プログラム評価に際し、プログラムの有効性に関する判定は不可欠であるという点では、合意しているのである。ただし、スクリヴェンが、教育のプログラム評価の場で RCT の実施が成功することはほとんどなく、現在では、介入プログラムの効果を決定する良質の調査デザインが他にもあると主張したのに対し、リップシーは同意せず、そのような他の方法に対しては懐疑的であった

とされる (Donaldson & Christie 2004)。

従来、アメリカの評価関係者の間では、定量的評価の優位性を主張する立場と、定性的評価を重視する立場との間で、論争が繰り返されてきたが、NCLB 法を機にこの論争が再燃したとされる (田辺 2006)。この論争は結論や合意が出たわけではなく、また、後述するように、RCT のみを優先して認めるという方針のために、WWC が行うエビデンスのレビューや評価に関して問題が指摘されるということも生じたのである。

3. WWC の概要

(1) トピック分野

WWC は教育の幅広い分野を扱うとされるが、何をレビューするか、に関する、トピック、介入、研究についての提案は、広く、保護者、教員、出版関係者、研究者、その他「何が有効か」の判定に関心をもつ個人や組織等、誰でも行うことができるとされる (Boruch & Herman 2007)。こういった点に、エビデンスの仲介機関として、関心を持つすべての人から質の高い教育研究とエビデンスに関する情報とニーズを収集しようとする WWC の姿勢が見られる。

トピックの候補に関する意見は、WWC が関わる専門家のフォーラムや、主要分野の専門家など、IES へのアドバイザーが提出することもある。WWC がレビューの対象とするトピックの選択は、

- (a) そのトピックの現行の教育政策と実践への関連性
- (b) どのような介入が採用されるべきか、という決定におけるトピックの見込まれる重要性
- (c) 存在するエビデンスのレベル

によって決定される、となっている。

WWC 設立初期のレビュー対象のトピック分野は、「ミドルスクールの数学」1 つだけであったが、2004 年から 2006 年の間に、「早期教育」、「初期読解力」、「小学校の算数」、「中退防止策」、「(英語を母語としない) 英語習得者」、「道徳教育」を加えた 7 つに増加した。その後、トピックは更に追加され (「青少年期リテラシー」「学習障害を持つ生徒」「高校の数学」「障害を持つ子どものための早期教育」)、2011 年 2 月現在ではトピックの数は 11 になっている。

WWC では、教育介入の効果について産出されるエビデンスの質を保証するために、エビデンス評価のための運用基準が開発、体系化され、可能な限り WWC のウェブサイト公開され、透明性を確保しているとされる (Boruch & Herman 2007)。

Boruch&Herman (2007) によれば、各トピックには、レビューチームが設置され、チームメンバーはプロジェクト代表者、プロジェクトコーディネーター、コーダーから構成される。各トピックの分野における WWC のレビューで、プロジェクト代表者はまず、詳細にわたるプロトコルを開発、記述するが、このプロトコルでは、介入、包括的な基準、リスクの高い下位グループを含む対象グループ、関連する従属変数、WWC 基準に基づいた WWC のレビューに適格な調査デザインとはどういうものか、などについての定義がなされる。

WWC におけるレビューのプロセスの最初の段階で、徹底した文献探索と、出版済み及び未出版の論文の全文の読解が行われる。この段階で、証言や単純相関のみに依拠している研究結果は除外され、RCT と精緻な準実験が WWC のレビューの候補として認められる。適格な研究だと確認されると、RCT と準実験デザインとの、基本的なカテゴリーにおける識別の作業が開始される。

設計や実施の段階で、重要な欠陥のない RCT は、WWC によって「エビデンス基準を満たす」と判定される。準実験で（１）プレテスト（または良質の代理テスト）と他の適切な対応変数において一致するものか、または（２）これらの変数において共変するものは、「留保付きでエビデンス基準を満たす」と判定される。「留保付き」という言葉は、全ての条件を同一にすると、RCT であれば、差異についてバイアスのない推定が確実にできるが、準実験ではそれができない、ということを意図的に示唆するものだと説明されている（Boruch & Herman 2007）。

研究のそれぞれのカテゴリについて、内部妥当性に影響を与える特徴が特定される。例えば、介入の実験毎の脱落率が大きく異なる RCT は特定される。そのような研究は、脱落が生じるバイアスについて説明する情報が不足している場合は「留保付きで基準を満たす」として、準実験のステータスまで評価を下げられる可能性がある（Boruch & Herman 2007）。

(2) 現在のトピック分野と「トピックレポート」¹⁰⁾

現在、WWC がレビューの対象としているトピック分野は以下のとおりである。各トピックの概要は、WWC の説明に基づいている。

■ 青少年期リテラシー（読み書き能力）（Adolescent literacy）

4 学年から 12 学年（高校 3 年生）までの青少年期リテラシー（読み書き能力）のカリキュラムと指導方法に関する調査研究についての評価が報告されている。これらのカリキュラムや指導方法は、アルファベット・読解の流暢さ、理解力のスキルや、一般的リテラシー（読み書き能力）のスキル向上を目的としている。

■ 初期読解力（Beginning Reading）

未就学児から 3 学年までの初期読解力のためのカリキュラムと指導方法に関する調査研究についての評価が報告されている。これらのカリキュラムや指導方法は、アルファベットや、読解の流暢さ、理解力のスキルや、一般的な読解達成度向上のためのスキル向上を目的としている。

■ 道徳教育（Character Education）

小学校、ミドルスクール、高校での活用のために設計された道徳教育のカリキュラムと指導方法に関する調査研究についての評価が報告されている。これらのカリキュラムや指導方法は、肯定的な人格の形成、好社会的な態度、学業成績などに関する生徒のアウトカム向上を目的としている。

■ 中退防止策（Dropout Prevention）

ミドルスクールと高校での、学校やコミュニティを基盤とした、中退防止のためのカリキュラムと指導方法に関する調査研究についての評価が報告されている。これらのカリキュラムや指導方法は、生徒が学校に在籍し続け、卒業する手助けとなるように設計され、学校で学習進歩を妨げる要因を緩和したりするようなサービスやアクティビティを含むこともある。

■ 早期教育（Early Childhood Education）

3 歳から 5 歳までの子どもが使うために設計されたカリキュラムや指導方法に関する調査研究についての評価が報告されている。これらのカリキュラムや指導方法は、言語、リテラシー（読み書き能力）、算数、認識力などを含む、就学前の学校レディネス（school readiness）に関連する能力開発を目的としている。

■ 小学校の算数（Elementary School Math）

未就学児から 5 学年までの、算数の学力に関わる生徒のアウトカム向上の目的で設計された、小学校の算数のカリキュラムや指導法に関する調査研究を評価した「介入レポート」を公表している。

■ 英語を母語としない英語習得者 (English Language Learners)

英語を母語としない、英語習得者のためのカリキュラムや指導法に関する調査研究を評価した「介入レポート」を公表している。これらのカリキュラムや指導方法は、未就学児から 6 学年までに焦点を絞り、読解力や算数のスキル向上や、英語力の発達を目的としている。

■ ミドルスクールの数学 (Middle School Math)

ミドルスクール (6 学年から 9 学年レベル) の数学の成績に関わる生徒のアウトカム向上のために設計されたカリキュラムや指導方法に関する調査研究を評価した「介入レポート」を公表している。

■ 学習障害を持つ生徒 (Students with Learning Disabilities)

K-12 (一般的に 5 歳から 18 歳) の学習障害を持つ生徒の学業成績向上を目的としたカリキュラムや指導方法に関する研究を評価した「介入レポート」を公表している。これらのカリキュラムや指導方法は、読解、算数／数学、書き、理科、社会科、学校での発達を含むアウトカム分野での学業成績の向上を目的としている。

■ 高校の数学 (High School Math)

9 学年から 12 学年の高校生の数学での介入に関する調査研究を評価した「介入レポート」を公表している。これらのカリキュラムや指導方法は、代数、幾何、三角法、微積分のスキルの向上を目的としたものである。

■ 障害を持つ子供のための早期教育 (Early Childhood Education for Children with Disabilities)

障害をもつ未就学児の学校レディネス向上を目的としたカリキュラムと指導方針に関する調査研究を評価した「介入レポート」を公表している。

「トピックレポート」は、トピック分野に関する介入についてのレポートからの情報を収集し、有効性、改善の指標、エフェクトサイズ¹¹⁾ についての評価についての概観を提供しており、その内容は以下のような項目に分かれている。

- 概観 (Overview)
- プログラムに関する追加情報 (Additional Program Information)
- 研究 (Research)
- 有効性 (Effectiveness)
- 参照文献等 (References)
- 資料 (Appendices)
- 「介入レポート」 (Intervention Report)
- 技術に関する資料 (Technical Appendices)

「トピックレポート」では、利用者が、あるトピックについての数多くのプログラムについて、評価や指標を簡単に比較できるようにしてある。

いったん公表されると、「トピックレポート」は更新されない。そのかわり、利用者はそれぞれのトピック分野で、研究や「介入レポート」(後述)の最新情報を基にして「研究知

見の要約を自分の関心に合わせて作成すること」を勧められる。

(3) 「トピックレポート」の例

■ 小学校の算数 (Elementary School Math)

WWC のウェブサイト (http://ies.ed.gov/ncee/wwc/reports/elementary_math/topic/) に記載されている内容によれば、この「トピックレポート」の知見は、2005-2006 年に行われた小学校の算数カリキュラムに関する WWC の初回報告の要約である。以下は、この「トピックレポート」の抜粋である。

本レポートでは、算数の知識とスキルを向上させるように設計された、小学校（幼稚園から 5 学年まで）の算数のカリキュラムについて調査している。

調査対象の算数カリキュラムは、数、計算、幾何学、代数（準備）、計測、グラフ、論理的証明のいずれかを教える教材を基盤とした指導プログラムで、再現可能なものである。

調査は、コアの算数カリキュラムだけを対象とした。コアの算数カリキュラムとは、1 学期もしくはそれ以上にわたる指導プログラムで、生徒の通常の学校の指導の中心となるもので、教科書、教具、コンピューターソフト、ビデオテープ、その他の教材の組み合わせを基本としている。このレビューでは、生徒の算数の学力達成を主要なアウトカムとして、焦点を絞っている。

まず、340 件の小学校の算数のカリキュラムの報告とされる研究を見たところ、介入に対する評価として、WWC のレビューの基準に適していたのは 237 件だった。残りの 103 件は介入のカテゴリーには属さなかった。

237 件の研究のうち、5 つのカリキュラムに関する 9 件の研究が、WWC 基準を満たしており、このうち 2 件が留保なしの適合、7 件が留保付きの適合であった。全体では、WWC は、73 のカリキュラムを見たわけだが、このうち 5 つは、留保なし又は留保付きで、WWC 基準を満たすものだった。67 のカリキュラムは、WWC のエビデンス基準を満たさなかった。1 つのカリキュラムは、単一事例の研究であり、これはレビューを保留されている。

5 つの小学校の算数カリキュラムの 1 つの領域について結果を見てみると、「Everyday Mathematics」というカリキュラムは、算数の成績にプラスの可能性の効果が認められたが、他の 4 つのカリキュラムは、算数の成績に明確な効果は見られなかった。

（以上、http://ies.ed.gov/ncee/wwc/reports/elementary_math/topic/より引用・抜粋。）

(4) 「介入レポート」 (Intervention Report)

あるトピックについての介入に関する研究のレビューは、継続的に、都度、更新され、公表されている。例えば、「小学校の算数」のトピックについての「トピックレポート」は上述のとおりだが、介入の算数プログラムの評価についてのレビューは随時行われており、2011 年 2 月現在では、73 件のプログラムについて、随時更新されたレビューの概要が公表されている。例えば、上述のトピックレポートで、唯一プラスの可能性の効果があると評価された「Everyday Mathematics」についての「介入レポート」は以下のような内容となっている。

<プログラムの概要>

「Everyday Mathematics」は、Wright Group / McGraw-Hill によって出版されている、幼稚園未就

園児から6学年までの生徒を対象としたカリキュラムで、概念と実践的なスキルを学習する多様な機会を提供している。各学年レベルにおいて、概念が様々な指導文脈のもとで復習され、広げられるようになっている。特に、このプログラムでは、日常生活での問題解決、算数的思考を生徒が伝えられること、テクノロジーの適切な活用に重点をおいている。また、共同学習を含む様々なタイプの指導法のバランスをとり、スキルの練習のための様々な方法を用い、生徒の学習に親が関わるように促すことも強調されている。

<調査>

「Everyday Mathematics」について「小学校の算数」のレビュー・プロトコルの範囲に適合した1件の調査が、WWCのエビデンス基準の「留保付き」に該当した。その調査にはテキサス都市部の大規模学区の3,436人の3学年から5学年までが含まれている。この学区では「Everyday Mathematics」の初版を使用した。

この調査に基づき、WWCは「Everyday Mathematics」の、小学生の算数の成績に対するエビデンスの程度は小さいと判断した。

<有効性>

「Everyday Mathematics」は小学生の算数の成績に対し、プラスの効果の可能性があるとされた。

	算数の成績
有効性の評価	プラスの可能性
改善指標	+11パーセンタイル順位

<プログラムの補足情報>

補足情報として以下の内容が簡単に紹介されている。

- ・開発者と連絡先
- ・利用の範囲（利用数等）
- ・指導（プログラムのゴール、補助教材、研修オプション等）
- ・費用

<調査>

WWCによってレビューされた72件の調査が、「Everyday Mathematics」の、小学校の生徒に対する効果を調査していた。1件の調査(Waite 2000)はWWC基準を留保つきで満たす準実験であった。残りの71件の調査は、WWCのエビデンス基準または適格性の選考で除外された。

Waite (2000) は、「Everyday Mathematics」を使用している6つの学校に在籍する732人の3、4、5学年の生徒と、算数の成績の点数、生徒の人口統計、地理的位置の基準値をマッチングさせた、12の同じような学校に在籍する2,704人の3、4、5学年の生徒から成る比較グループを調査した。介入グループの学校は、「Everyday Mathematics」の初版を実施する最初の年であった。比較グループは、学区によって承認された伝統的な算数カリキュラムを使用した。

・エビデンスの程度

WWCはそれぞれの領域で、エビデンスの程度を小さい、中程度、大きい、に分類している（詳細はWWCの「手続きと基準ハンドブック」に記載）。エビデンスの程度は、調査の数とWWC基準

に留保なし或いは留保つきで適合した調査全体でのサンプルサイズの総数を考慮している。

WWC は、小学生に対する「Everyday Mathematics」のエビデンスの程度は、小さいと考えている。

<有効性>

・調査結果

小学校の算数の介入についての WWC のレビューは、算数の成績における生徒のアウトカムに焦点を当てている。下記の調査結果は、「Everyday Mathematics」の小学生に対する効果（エフェクト）のサイズと統計的有意性について、調査を行った著者の推定と、WWC が計算した推定が含まれている。

Waite (2000) は、算数の成績全体に対する「Everyday Mathematics」の統計的に有意なプラスの効果を報告した。WWC の計算では、この効果は、統計的に有意ではなかった。しかし、WWC は、算数の成績に対する効果は実質的に重要である（すなわち、エフェクトサイズが 0.25 以上である）と判定した。この 1 件の調査に基づき、WWC は、算数の成績全体に対する「Everyday Mathematics」の効果は実質的に重要なプラスの効果がある、と判定した。

Waite (2000) はまた、下位検査（概念、演算、問題解決）の結果についても統計的に有意なプラスの効果を報告している。WWC の計算では、各下位検査の効果は統計的に有意ではなかった。しかし、各下位検査の効果は、実質的に重要だと考えるに十分大きかった。下位検査の分析は全体のサンプルの結果の一部として既に表されているので、介入についての有効性の評価には盛り込まれない。

・有効性の評価

WWC は、アウトカムの分野における介入の有効性をプラス（positive）、混在（mixed）、明確な効果なし（no discernible effects）、マイナスの可能性（potentially negative）、マイナス（negative）で評価している。有効性の評価は 4 つの要因：調査デザインの質、調査結果の統計的有意性、介入と比較の条件の参加者の間の差のサイズ、調査全部を通した結果の一貫性、を考慮している（詳細は WWC の「手続きと基準ハンドブック」に記載）。

・改善指標

WWC は個々の調査結果について改善指標を算定している。さらに、各アウトカムの領域内で、WWC は、各調査の平均改善指標と、調査全部にわたる平均改善指標を算定している（詳細は WWC の「手続きと基準ハンドブック」に記載）。改善指標は、介入条件での平均の生徒のパーセンタイル順位と、比較条件での平均の生徒のパーセンタイル順位の差を表したものである。有効性の評価と異なり、改善指標は有効性の統計的優位性や調査デザイン、あるいは分析に関わらず、完全にエフェクトのサイズに基づいている。改善指標は-50 から+50 の間の数値をとり、プラスの数字は介入グループに有利な結果を示している。

WWC 基準を留保つきで満たす 1 件の調査に基づき、算数の成績に対する改善指標は+12 パーセンタイル順位である。

・要約

WWC は小学生への「Everyday Mathematics」の 72 件の調査をレビューした。このうちの 1 件の

調査は WWC 基準を留保つきで満たし、残りの 71 件の調査は、WWC のエビデンス基準または適格性に適合しなかった。この調査に基づき、WWC は、小学生の算数の成績全体に対する「Everyday Mathematics」の効果はプラスの効果の可能性がある、と判定した。この報告書に提示された結論は、新しい調査が出現するにつれて変わることがある。

<参考文献>

・ WWC エビデンス基準に留保つきで適合

Waite,R.D. (2000). A Study of the effects of *Everyday Mathematics* on student achievement of third-, fourth-, and fifth-grade students in a large north Texas urban school district. *Dissertation Abstracts International*,61(10), 3933A.(UMI No.9992659)

・ WWC のエビデンス基準または適格性に適合しなかった調査

(以下省略。以上、http://ies.ed.gov/ncee/wwc/reports/elementary_math/eday_math/index.asp より引用。)

この「介入レポート」でも、先述の「トピックレポート」でも、WWC のエビデンス基準に適合する調査が非常に少ないことが顕著である。1 件しか WWC のエビデンス基準を満たさない評価に対して、その厳密性に対する信頼がどの程度のものであるのか、推測が困難である。WWC がそのエビデンス基準を固守する限り、基準を満たす研究が増えるのを待つほかないわけだが、実際のところ、教育省の支援によって、教育分野での RCT による調査研究は増加傾向にあるとされる (Slavin 2008)。したがって、こういった過程を通じ、WWC は、エビデンスの仲介機関として、科学的に根拠のある調査研究を推進するという機能を果たしているということになるのであろう。

(5) 「実践ガイド」(Practice Guide)

「実践ガイド」は、教員が、教室や学校で直面する日常の課題に対処する助けとなるための、実務的な提案を提供しており、2011 年 2 月現在で 14 件のガイドが発表されている。

このガイドは、全国的に著名な専門家で構成される委員会によって作成され、その内容には、実行可能な提案、ありうる障害を乗り越えるための方法、そしてそれぞれの提案をサポートするエビデンスの強さの表示が含まれる。この IES「実践ガイド」は、厳密な外部ピアレビューを受けることになっている。

現在公表されている「実践ガイド」

- 「幼稚園から 8 年生までの効果的な分数の指導を開発するには」(Developing Effective Fraction Instructions for Kindergarten through 8th Grade) (2010 年 9 月発表)
- 「幼稚園から 3 年生までの読解力を向上させるには」(Improving Reading Comprehension in Kindergarten through 3rd Grade) (2010 年 9 月発表)
- 「生徒の学業成績データを指導上の意思決定に活用するには」(Using Student Achievement Data to Support Instructional Decision Making) (2009 年 9 月発表)
- 「大学までの道のりをサポートするには: 高校には何ができるか」(Helping Students Navigate the Path to College: What High School Can Do) (2009 年 9 月発表)
- 「学校外の時間を学力向上に向けて組み立てるには」(Structuring Out-of-School Time to Improve

Academic Achievement) (2009 年 7 月発表)

- 「数学が苦手な生徒の支援：小学校、ミドルスクールでの介入に対する反応」(Assisting Students Struggling with Mathematics: Response to Intervention for Elementary and Middle Schools) (2009 年 4 月発表)
- 「読解が苦手な生徒の支援：低学年での介入及び多層介入に対する反応」(Assisting Students Struggling with Reading: Response to Intervention and Multi-Tier Intervention in the Primary Grades) (2009 年 2 月発表)
- 「小学校の教室での問題行動を減少させるには」(Reducing Behavior Problems in the Elementary School Classroom) (2008 年 9 月発表)
- 「中退防止策」(Dropout Prevention) (2008 年 8 月発表)
- 「青少年期のリテラシー：効果的な授業と介入実践」(Improving Adolescent Literacy: Effective Classroom and Intervention Practices) (2008 年 8 月発表)
- 「慢性的に成績の低い学校を劇的に変えるには」(Turning Around Chronically Low-Performing Schools) (2008 年 5 月発表)
- 「小学生の英語習得者のためのリテラシーと英語の効果的な指導法」(Effective Literacy and English Language Instruction for English Learners in the Elementary Grades) (2007 年 12 月発表)
- 「数学と理科で女子を奨励するには」(Encouraging Girls in Math and Science) (2007 年 9 月発表)
- 「生徒の学習を向上させる指導法と研究を体系づけるには」(Organizing Instruction and Study to Improve Student Learning) (2007 年 9 月発表)

(6) 「実践ガイド」の例

- 「大学までの道のりをサポートするには：高校には何ができるか」(Helping Students Navigate the Path to College: What High School Can Do)(2009 年 9 月発表)

以下は、WWC による、「大学までの道のりをサポートするには：高校には何ができるか」という「実践ガイド」から一部を抜粋したものである。

1988 年から 2008 年までの、米国内の大学進学準備プログラムについての研究は、500 件以上存在する。このうち、99 件の研究が、因果関係を調べる目的で設計され、中等学校レベル（6 学年から 12 学年まで）を、大学に向けて学力面で準備させ、進学申請等の手続きをサポートし、大学入学の可能性を高めることを目的としている。

WWC 基準に従ってレビューを行ったところ、10 のプログラムについての 16 件の研究が WWC 基準を「留保なし」または「留保付き」で満たしていた。

大学進学プログラムは、生徒が大学に向けて準備し、入学するまでにしなければならない様々なステップを扱う多数の部分から構成されている。

複数の実践が一つのプログラムにまとめられていると、一つずつの実践の効果の評価が難しくなる。これは「実践ガイド」作成にとっては問題である。「実践ガイド」は、学校が大学への進学向上のために実施できる具体的な方法を提供しようとしているからである。エビデンスを基に委員会では、下記のような提案を行った。

なお、多くの場合、一つのプログラムが、「実践ガイド」を作成した WWC の委員会の提案の一つの側面しか実施していない。例えば、研究が基準を満たしている、いくつかのプログラムは、提案 3 の最初のステップの「メンタープログラム」は実施しているが、ステップ 2 の「ピアグループ」とステップ 3 の「キャリア探索」は実施していない。また、多くのプログラムが実施している実践は、委員会の提案と完全にそろっているわけではない。提案されている実践をプログラムがどの程度実施しているのか、研究報告では詳細が不足していることもある。

委員会は、基準を満たしていた研究の対象のプログラムの実施報告についてレビューを行い、それぞれのプログラムの構成部分の相対的な重要度を評価し、プログラム実施についてより深く理解しようとした。そして、以下の点を考慮した後、各提案にそれぞれエビデンスのレベルを付与した。それらは、(1) 基準を満たす研究対象のプログラムの数をプログラムが提案を実施している度合い、(2) プログラムが、高校の学業成績に与える影響、(3) 大学入学申請の重要なステップの修了、(4) 大学進学、である。

提案 1：生徒が大学レベルの勉強の準備ができるようなコースとカリキュラムを提供し、大学への準備ができているカリキュラムとはどういうものか、生徒に 9 学年までに確実に理解させる。
(エビデンスのレベル：低い)

提案 2：生徒が大学に進学する準備がどの程度できているか気づくように、高校時代を通して評価指標を活用し、足りないところがあれば、それが明らかになるように生徒を支援する。(エビデンスのレベル：低い)

提案 3：生徒の周りに彼らの大学へ行きたいという希望を育て、サポートしてくれるような大人と友人を置く。(エビデンスのレベル：低い)

提案 4：大学に入るための重要なステップを生徒が完全に踏めるように励まし、支える。(エビデンスのレベル：普通)

提案 5：大学に入るための経済的な必要性について家族の認識を高め、生徒が奨学金に応募するのを手伝う。(エビデンスのレベル：普通)

(以上、http://ies.ed.gov/ncee/wwc/pdf/practiceguides/higher_ed_pg_091509.pdf より引用。)

このように、「実践ガイド」は、特定のプログラムの評価ではなく、プログラムの評価を行った結果のエビデンスを基に、委員会が具体的な提案を行っているものである。

(7) 「クイックレビュー (Quick Review)」

2008 年に新しく始まったサービスで、主要な全国ニュースメディアで新しく公表された研究論文や研究報告に関して、教育関係者に、研究の質に関して時宜を得た客観的な評価を提供することを目的としている。

このレビューでは、幼稚園就園前から高等教育レベルまで、教育や学校を基盤とした介入の有効性に関する研究に焦点を絞っている。

例えば、2008 年 6 月に発表された「クイックレビュー」では、「6 学年でミドルスクールを始めるのはマイナス効果」(Cook, P., J., McCoun, R., Muschkin, C., & Vigdor, J. (2008) Journal of Policy Analysis and Management, 27(1), 104-121.) という調査について、以下の内容が 1 ページにまとめられている。

<この調査は何についてのものか？>

この調査は6年生を小学校ではなくミドルスクールに入れることで行動や学業成績に影響が出るかどうか、を調べたものである。調査の分析対象となったのは、2000-2001年の期間に、ミドルスクールに通った約4万人の6年生と、小学校に通った約5千人の6年生に関するノースカロライナのデータである。調査の著者らは分析に際し、統計的マッチングを用いて似たようなミドルスクールと小学校を選んだ。人種的・民族的構成、テスト平均点、その他の特徴をマッチングさせた。著者らは、全ての公立学校についての州全体のデータベースに基づくテストの点数と規律違反のデータを分析した。

<6年生のどんな2つのグループが比較されたのか？>

一方のグループは、6、7、8年生が通うミドルスクール（6年生は平均257人在籍）に通い、もう一方のグループはK-6（未就学児～6年生）の小学校（6年生は平均117人在籍）に通った。

<調査の著者らは何を報告しているか？>

ミドルスクールに通った6年生は小学校に通った6年生に比べ、規律違反の数が2倍であった。ミドルスクールの6年生はまた、標準化の読解テストが、小学校の6年生に比べて低い点数だった。算数のテストでは、統計的に有意な違いは見られなかった。

調査結果は、分析で統制されていない調査グループの間の違いに影響を受けている可能性があるため、WWCは結果については留保している。

<WWCの評価>

この記事で説明されている調査は、WWCのエビデンス基準を留保付きで満たしている。

- ・この調査の長所

ミドルスクールと小学校をマッチングさせるために統計的手続きを使用した。

- ・注意

統計的マッチングにもかかわらず、2つのグループの学校群の特徴にいくつかの違いが残った。例えば、規模の大きい学区のミドルスクールには、経済的に不利な生徒が少ない。著者らは、これらの差異を調整するために、統計技術を用いた。しかし、学区の方針や生徒集団の特徴における違いや、分析で統制されていないその他の要因が結果に影響を与えている可能性がある。

（以上、<http://ies.ed.gov/ncee/wwc/publications/quickreviews/QRReport.aspx?QRID=12>より引用。）

尚、注意書きとして、WWCは、『クイックレビュー』は引用した文献上で発表されたエビデンスに基づき、調査の著者らの報告するエフェクトサイズと有意レベルに頼っている。WWCは、著者らの調査結果を確認したり、調査に関する追加情報のために著者らに連絡をとったりしていない。WWCの評価は、上記の文献に要約された結果のみを参照しており、必ずしも調査で提示された全ての結果を参照しているとは限らない。」と、説明している。

「クイックレビュー」で扱っている調査研究は一覧できるようになっており、2011年2月現在で55件のレビューが載せられている。扱われる介入のタイプは多岐にわたり、その内容と件数はそれぞれ「カリキュラムと指導法」（14件）「学業補助」（5件）「学校組織と運営（ガバナンス）」（14件）

「生徒に与えるインセンティブ」(6件)「教員プログラム」(9件)「生徒の行動」(7件)「オールタナティブ教育プログラム」(1件)「その他の介入」(2件)となっている。(55件のうち3件は2つのタイプに分類されている。) ¹²⁾ この内訳を見ると、「クイックレビュー」は、かなり幅広い分野や実践に及んでおり、WWC が意欲的に教育の分野における広範囲での調査研究についてレビューを行っていることが推察される。

(8) マルチメディア (Multimedia)

WWC は、教育に関する意思決定でのエビデンス活用に関するセミナーやパネルディスカッション、ウェブ上のセミナー (Webinar) などを主催している。このイベントの内容は、音声ファイル、プレゼンテーション、筆記録などの形式で、ダウンロードが可能である ¹³⁾。

4. WWC の課題

(1) RCT 最優先に関わる問題

エビデンスの仲介機関として WWC が行っているエビデンスの統合・評価には、その性質上、様々な問題が伴うとされるが、中でも、WWC が RCT のみを重要視することによって問題が生じているという指摘がなされている (Slavin 2008)。

例えば、RCT であればよし、とするあまり、WWC の過去のレビューでは、サンプルサイズが非常に小さいエビデンスや、実験の方法・内容に問題があるエビデンスが高い評価を受けている例があるとされる (同上)。

Slavin は、多くの教育の介入プログラムの調査研究では、サンプルサイズが非常に小さい例が多いと指摘した上で、調査の「生徒」の数が少ない場合、明らかに統計的な信頼性 (統計検出力) に問題が生じるが、更に「クラス」や「学校」の数が少なすぎると、教員やクラスの特徴が大きく影響する可能性があり (交絡因数)、この数が大きければ大きいほど、影響は小さくできると説明している。

更に、小さいサンプルサイズの調査は、肯定的にも否定的にも極端なエフェクトサイズを出す傾向があり、学校や教員、クラスの効果は、大規模な調査では差が小さいが、小規模の調査では大きい影響を与えうると指摘している。

また、大規模な調査はたとえ、実験の結果がゼロや否定的であっても、実験にかかったコストや背景から、結果を出版しないという選択は難しいが、小さい実験では、結果がゼロや否定的であれば、発表されない可能性が大きい。したがって、小規模の実験は結果がプラスのものが発表されやすい、という「出版バイアス」も存在する、とされる (Slavin 2008)。

Slavin は、小規模の実験に必ず問題があるという訳ではなく、小規模の実験の集合体は、こういった出版バイアスによって、否定的な結果の研究が研究者の引き出しに眠っている可能性がある「お蔵入り効果」 (file drawer effects) 等の問題があるため、大量の研究のメタアナリシスでは注意が必要だと説明している ¹⁴⁾ (2008)。

RCT はサンプルサイズが小さくなりがちであり、特に個々の生徒がランダムに割り当てられる場合にはそうなる傾向がある。こういった場合、上述のようなバイアスが生じやすくなる。大規模のマッチングの調査のほうが、小規模の RCT よりも意義があり、信頼できる情報を提供できる可能性もあるとされる。 (Slavin 2008)。

しかし、WWC は、RCT の対象となる各グループに 1 人の教員や 1 つの学校しか含まれない調査は除外するが、サンプルサイズについては制限を設けていない。結果的に、WWC による評価では、非常に規模の小さい（RCT による）調査が、多くの介入プログラムの評価を決定することになる（Slavin 2008）。

例えば、過去の「ミドルスクールの数学」のトピック分野でのレビューで「Saxon Math」と呼ばれるプログラムがプラスの効果があるという高い評価が出されたが、この根拠となった唯一の調査は、わずか 46 人の生徒を対象としたものであり、しかも比較に使われたテストは「Saxon Math」で学習した生徒しか学んでいない内容についてのものであった（Slavin 2008）。（尚、この「Saxon Math」については、WWC はレビューとそれに伴う「介入レポート」を 2010 年 2 月に更新し、新しいレポートでは、Slavin が問題としていた調査については、該当学年が不適当として評価不可、とされている 15)。）

また、「英語を母語としない英語習得者」のトピック分野で WWC レビューの高い評価を受けた「Peer Tutoring and Response Group（級友指導と応答グループ）」の実践では当初、評価の根拠とされたのは、46 人の生徒を対象に行われた調査であったと指摘している。

更に、WWC の初期読解力プログラムのレビューでは、Daisy Quest という、フォニックス（つづり字と発音の関係を教える語学教授法）のプログラムが高い評価を受けているが、このレビューで検証されたエビデンスは、実験時間がわずか 5 時間未満であり、しかも研究チームがこのプログラムを実際に子供達に教え、さらに対象グループと比較グループとの比較テストでは、比較グループの生徒には全く内容が教えられていないフォニックスについてであった、とされる（Slavin, 2008）。

Slavin（2008）は、WWC のレビューと評価が、レビューの対象が RCT であることのみに注意を払いすぎるあまり、サンプルサイズやプログラムの介入期間の長さ、その他、非論理的な結論にたどり着く可能性のあるバイアスを招く恐れのある実験デザインの要素を軽視している、と批判している。

そもそも、RCT は、特に教育の場では大きいサンプルサイズをとることが困難である。しかもその上、非常に限られた人数と設定で行った RCT によるプログラムに高い評価を与えられたとしても、そのプログラムを実際の教育の現場で実施することは、現実的に無理があり、たとえ、そのような結果を応用したとしても、どこまで同じ結果が得られるのか、といった疑問が残るとされる（Slavin 2008）。

こういった点からも、Slavin は、エビデンスの統合評価では、RCT だけを絶対視するのではなく、適切なサンプルサイズによってマッチングを行った準実験についても相当の評価をするべきだとしている。

実際、米国の教育の分野には現在、WWC 以外にもいくつかの統合評価を行う the Comprehensive School Reform Center や the Best Evidence Encyclopedia などの機関や、Social Programs that Work (www.evidencebasedprograms.org) や Promising Practices Network (www.promisingpractices.net) などのウェブサイトが存在し、それらの中には必ずしも RCT のみを最優先とはせず、良質の準実験も採用するなど、より合理的なエビデンスの選択を行う場合もあるようである（Slavin 2008）。エビデンスの仲介機関が WWC だけではなく、複数存在することによって、利用者側は、提供されるエビデンスを比較評価できる選択が与えられることになり、そのことによってもたらされる健全性の意義は大きいと考えられる。

(2) 現行の教育改革における WWC の課題

(a) 2つのパラダイムの衝突

冒頭で述べたように、米国では 1990 年代以降、各州でスタンダードが開発されてきたが、スタンダード教育改革について二つのパラダイムが混在している、という指摘がなされている（松尾 2010）。すなわち、二つのパラダイムとは行動主義パラダイムと構成主義パラダイムという、異なる考え方の学習理論である。

20 世紀を通して中心的な学習理論とは行動主義パラダイムであり、その考え方によれば、学習とは、条件づけによる訓練によって行動が変容することを意味した（松尾 2010, Shoenfeld 2006）。行動主義に基づく教育実践では、スタンダードは教えられるべき知識や技能が細分化されて定義され、ドリル学習や知識の記憶と再生が繰り返され、標準テストによって知識や技能の習得状況が測定される（松尾 2010）。

これに対し、1970 年代から 1980 年代にかけての認知科学の進展によって、構成主義による学習理論が発達し、学習は、生徒が持っている認知の枠組みに新しい知識を関連づけ、組み換える行為だとされた。構成主義パラダイムに基づく教育実践では、知識を構成する主体者は生徒だという認識に立ち、生徒自らが問題解決のプロセスを経験できるように、教員は学習の方向づけを行うとされる（松尾 2010）。

この二つのパラダイムの衝突は、教育の様々な分野で論争を引き起こしたが、ここでは、数学のカリキュラムをめぐる議論を、WWC との関わりから見ることにする。

米国の教育におけるカリキュラムの中でも、特に数学のカリキュラムは、国家の経済的、軍事的な発展にかかわる問題として、歴史的に論争のもととなり、ときに大きく修正、変更の途を辿ってきた（Shoenfeld 2004）。加えて、1980 年代後半には、the Second International Mathematics Study という国際比較テストによる米国の生徒の成績の低さが問題となり、新しいカリキュラムを要求する声が高まるようになった（同上）。

Shoenfeld（2006）によれば、数学に対する、従来の行動主義パラダイムに則った考え方は「学習は知識の積み重ねであり、訓練によって熟達する。そして知識は研究され尽くした特定の種類の問題を解く能力によって証明される」というものであった。

ところが、構成主義パラダイムによって、効果的な数学の成績にはより幅広いスキルと理解が不可欠であることが実証されるようになり、従来のカリキュラムで重視されてきた、堅固な知識基盤を持つことの他に、幅広い問題解決の方法を使うことができること、数学的な考えを使って効果的に論理的な思考ができ、その論理を効果的に話したり書いたりして説明できること、知識と時間を含む、自分の自由になる資源をうまく使えること、数学の世界に対して生産的な信念と習性を持っていること、などが必要だとされるようになったのである。

これを受けて、National Research of Council（全米研究協議会）は数学の習熟を構成する柱は、以下の 5 つだとした。

- ・ 概念的理解：数学的な概念、操作、関係の理解
- ・ 手続きの流暢さ：手続きを柔軟に正確に効率的に適切に実行できるスキル
- ・ 戦略的能力：数学の問題を方針を立て、説明し、解決する能力
- ・ 適応的推論：論理的思考、内省、説明、正当化を行う能力

- ・ 生産的習性：数学を理にかなった、便利で価値のあるものとみなす習慣的傾向と、努力と自分の能力に対する信念

こういった動向を受けて、National Council of Teachers of Mathematics (NCTM) (全米数学教師協議会) は、1989 年、数学の問題解決のプロセスを重視する「スタンダード (the Standards)」を提唱し、スタンダード開発のさきがけとなった (Shoenfeld 2006)。

このスタンダードは、数学の学習における暗記やドリルよりも、問題を解く上での積極的な学習思考プロセスを重視し、それらを「プロジェクトを通じた学習」「グループと個人における作業」「教師と生徒、また生徒同士のディスカッション」などを通した指導によって学習されるべきだとした (Shoenfeld 2004)。それをどのように実践の場で行うか、については、個々の学校、教員の裁量に任されることになり、その結果、様々な「スタンダードに基づいた教育」が行われることになった。

この新しい「スタンダードに基づいた教育」では、問題解決の思考やプロセスを重視するため、従来のカリキュラムに比べ、基礎的スキルの習得に時間がかかる。従来のカリキュラムの支持者は、基礎知識は応用のために不可欠であり、この新しい「スタンダード」では、基礎的スキルの欠如によって生徒が深刻な被害を受ける、と批判した。一方、スタンダードを推進する人々は、スキルは、意味のある問題解決の文脈で発達すれば、より強固に発達する、と反論した (Shoenfeld 2006)。

Shoenfeld は、こういった、数学教育のカリキュラムについて「基礎重視」と「プロセス重視」の間で論争が続いている背景を説明した上で、どちらがより効果的なのか、本当の意味での比較は過去になされておらず、実のところ、この議論の存在こそが WWC 設立の動機の一つであった、と述べている (Shoenfeld 2006)。

「スタンダード」の考え方に基づく新しいカリキュラムでは、上述のように、従来は扱われなかった能力の学習、発達も目指されているが、従来のカリキュラムに対応する既存のテストでは、新しいカリキュラムで重視される能力がどの程度身についたか、は評価されないことになる。Shoenfeld (2006) は、実際、同じ生徒たちが、従来のスキル重視の標準学力検査 (この場合は the California STAR test) と、より幅広い能力を見る「スタンダード」に基づくテスト (the Balanced Assessment test) を受け、その成績を比較したところ、22%の生徒が標準学力検査では「熟達レベル」という評価を受けながら、「スタンダード」に基づくテストでは、「熟達レベルに達していない」という評価を受けたことを指摘し、標準学力検査では、評価される能力がせまい定義をされているために、評価が正確性に欠ける可能性がある、と指摘した。

また、2つのカリキュラム A と B を想定した場合、カリキュラム B よりもカリキュラム A から多く学んだ生徒が、A と B を比較する目的のための標準学力検査で (特に、スキル中心の RCT で) カリキュラム B 以上のことを学習した部分については何もテストされなかったら、標準検査が、2つのカリキュラムの重要な違いである、概念の理解や問題解決に無関心である以上、カリキュラム A と B の間に「有意な差は見られない」ことになる (Shoenfeld 2006)。

Shoenfeld は、こういった問題に WWC は積極的に取り組むべきだと提言したが、にもかかわらず、実際は、WWC のレビュー報告には深刻な問題が生じたと述べている。

例えば、WWC が行った初期の「ミドルスクールの数学」のレビューで、「the Expert Mathematician」というプログラムに関して、唯一 WWC 基準に合致した調査では、このプログラムにも比較のプログラムにも、同様のプラス効果が見られた。WWC はこの結果について詳しい説明をせずに、『the Expert Mathematician』は「効果がない」という評価を出したという (Shoenfeld 2006) ¹⁶⁾。

また、Shoenfeld (2006) は、同じ「ミドルスクールの数学」のレビューで、ある調査研究が Florida Assessment Test を使用しているために、統計上の信頼性と妥当性が確認できた、という理由で、WWC が高い評価を出している例を挙げ、Florida Assessment Test 自体が数学のどの能力をテストしているか、というテストの内容分析が必要であるにもかかわらず、WWC がその問題に関心であり、WWC 設立の目的を損ねている、と批判している¹⁷⁾。

Shoenfeld による上述の批判に対し、WWC は、WWC がレビューを行うテストの内容に何が含まれているか、一つ一つ検証することは時間と労力がかかり過ぎることになり、WWC 本来の役割がおろそかになるため、その作業は行わないことにしたのだ、と説明している (Herman 他 2006)。

上記の例に見られるように、エビデンスの仲介機関として、対立する2つのパラダイムについて、エビデンスの実証分析を通して検証や調整を行うことは、その必要性が指摘されているにもかかわらず、現行の WWC には荷が過ぎるとして予定されていないということであろう。

(b) アカウンタビリティにおける WWC の影響

米国の現行の教育改革では、NCLB 法によって、州や地方学区レベル毎に、生徒の学力の単年毎の向上達成度 (adequate yearly progress, 略称 AYP) が標準テストで判断されるため、このテストの点数の結果が絶対視され、州や学区ではそれに基づくアカウンタビリティが連邦政府によって要求されるようになった (杉浦 2006, 松尾 2010)¹⁸⁾。結果的に、標準テストとそれによって測定される学力のみが重視される教育につながっている、という批判がなされている (松尾 2010)。

このような、標準テストで AYP が求められる現在のアカウンタビリティ制度では、テストの点数にすぐ効果が現れる、読解や算数・数学の「効果的な」教育プログラムの導入に傾く方向にあるとされる。そして当然のことながら、教員による指導や生徒の学習も、それら特定の教育プログラムに合わせたものとなっていくと指摘されている (矢野 2003)。

連邦教育省の意向を受け、IES と WWC によって有効だと認められ、結果として広く採用されている教育プログラムは、標準テストの上では一定の学力向上につながっているといえるかもしれないが、それらは、教員の努力工夫によるカリキュラムの研究や開発をあまり要求しないものであり、教員の指導力の向上にはつながらない可能性も高いという見方がある (矢野 2003)。また、「効果的」と公に認知される教育プログラムに過度に依存することによって、教える側も学ぶ側も基礎的な技能の習熟に埋没してしまうのではないかと、という懸念も指摘されている (矢野 2003)。

スタンダード教育改革で目指されているのは、読み書きや算数などの基礎的な知識やスキルの習得ではなく、「グローバルな経済の中で勝ち抜くことのできる高度な思考力や判断力の育成」だとされる (松尾 2010)。にもかかわらず、スタンダードに対する2つのパラダイムの衝突や、アカウンタビリティを要求する NCLB 法によって、米国の教育改革が現実に進む方向は、必ずしも明確ではないように思われる。

スタンダードとアカウンタビリティとがセットになって推し進められている現行の米国の教育改革の下で、WWC は、エビデンスの仲介機関として、厳密なエビデンス基準を設定し、それに基づくエビデンスの収集、レビュー、評価を行い、できるだけわかりやすい形で利用者に報告を行っている。また、先に述べたように、WWC のエビデンス基準の普及も、WWC の閲覧件数の増加も連邦教育省によって推進されている。

こうした意味で、WWC は教育省が推進する「科学的に根拠のある」教育評価の基準を示し、そ

の普及の一翼を担うという、エビデンスの仲介機関としての機能を果たしつつあるといえるかもしれない。ただ、そのことが結果的に、現行の教育改革で目指されているはずの生徒の高度な思考力や判断力の育成につながっているのか、それとも基礎的な知識やスキルの習熟に向かうような、既成の「効果的」なプログラムの重視・推進で終わっているのか、という点については、さらなる検証が待たれる。

WWC の評価の対象は特定の学習プログラムだけでなく、「クイックレビュー」の対象は、広く学校運営などの実践にも及んでおり、さらに「実践ガイド」ではプログラムのエビデンス評価を基にした具体的な提案もなされているが、主な評価の対象はやはり様々な介入プログラムであろう。したがって、NCLB 法の今後の展開によっては、エビデンスの仲介機関としての WWC の役割が変わる可能性も否定できない。

NCLB 法によって、学力格差の是正が連邦政府の支援と圧力によって強力に推進され、エビデンスの仲介機関としての WWC の重要性が拡大される一方で、米国の現行の教育改革が向かっている先を、今一度見定める必要があるように思われる。そして、そのことは、我が国の教育におけるエビデンスのありかたや、新しい知と学力の形成を考える上でも重要な示唆を含むであろう。

注

- 1) <http://ies.ed.gov/director/>
- 2) <http://www.edweek.org/ew/articles/2007/07/18/43whatworks.h26.html>
- 3) http://www.mathematica-mpr.com/About_Us/
- 4) <http://www.edweek.org/ew/articles/2007/07/18/43whatworks.h26.html>
- 5) <http://www.edweek.org/ew/articles/2008/06/11/41whatworks.h27.html>
- 6) http://coalition4evidence.org/wordpress/?page_id=468
- 7) エビデンスを系統的にレビューするという考え方は、保健医療分野でのコクラン共同計画や社会科学のキャンベル共同計画によって 1990 年代から開始されており、教育分野でのエビデンスの系統的レビューを行う WWC の設立に大きい影響を与えたとされる（田辺 2006）。
- 8) <http://www.edweek.org/ew/articles/2008/06/11/41whatworks.h27.html>
- 9) 後者の意見を提出した中には、WWC と関係の深いボルーチ（Robert F.Boruch）やクック（Thomas D.Cook）らも含まれている。
- 10) <http://ies.ed.gov/ncee/wwc/reports/>
- 11) それぞれの研究においてアウトカムの測定スケールが異なるので、平均値や標準偏差の絶対値は異なるが、平均値の差を標準偏差で割り算した値は、1 標準偏差あたりの平均値の差を表すことになり、標準化された平均値の差を表すことになる。すなわち、測定のスケールが異なっても互いに比較することが可能な値になる。これを（標準化）エフェクトサイズと呼ぶ。（Pope 他 2007 より引用）
- 12) <http://ies.ed.gov/ncee/wwc/publications/quickreviews/>
- 13) <http://ies.ed.gov/ncee/wwc/publications/multimedia/>
- 14) 統計的にそれらを推定する措置がとられることもあるが、そういったケースは稀であるとされる（Slavin 2008）。
- 15) http://ies.ed.gov/ncee/wwc/pdf/wwc_saxonmath_042010.pdf
- 16) WWC は、この調査研究を 2006 年に再レビューし、評価の内容を「潜在的にプラスの効果が認められる」と修正したようである（http://ies.ed.gov/ncee/wwc/reports/middle_math/expert_math/index.asp）。
- 17) 尚、これらの問題等を議論した彼や他の研究者の報告書が WWC の特別版として出版される予定であったが、IES によって中止されたことも、当局による深刻な政治的圧力の問題だとして、Shoenfeld は批判している。
- 18) NCLB 法に基づく連邦政府の役割及び州政府や地方機関への強力な介入権限は先例がない（杉浦 2006、吉良 2009）上、法的に、州政府に財源が保障せずに介入することが禁じられている連邦政府の役割を超えているという批判もなされている。

引用・参考文献

- OECD 教育研究革新センター編著（岩崎久美子他訳） 2009『教育とエビデンスー研究と政策の協同に向けて』明石書店
- 荻谷剛彦 2001『教育改革の幻想』ちくま新書 筑摩書房
- 北野秋男 編著 2009『現代アメリカの教育アセスメント行政の展開ーマサチューセッツ州（MCAS テスト）を中心にー』東信堂
- 吉良直 2009「どの子も置き去りにしない（NCLB）法に関する研究ー米国連邦教育法の制定背景と特殊性に着目してー」教育総合研究：日本教育大学院大学紀要 2, 55-71, 日本教育大学院大学
- 現代アメリカ教育研究会（代表：金子忠史）1998『カリキュラム開発をめざすアメリカの挑戦』教育開発研究所
- 佐々木亮 2009「第2章 アメリカにおけるメタ評価の現状」「諸外国における政策評価のチェックシステムに関する調査研究ー報告書ー」2009 総務省行政評価局
- 佐々木亮 2010「エビデンスに基づく開発援助評価ー援助評価の歴史、ランダム化実験の起源、スクリヴェンとバナージェの考え方の比較ー」、『日本評価研究』10（1）pp.63-73
- 杉浦慶子 2006「米国ワシントン州の教育スタンダードに基づく学力政策ー制度的枠組みの生成と No Child Left Behind（NCLB）法の影響ー」東北大学大学院教育学研究科研究年報 第55集第1号
- 田辺智子 2006「エビデンスに基づく教育ーアメリカの教育改革と What Works Clearinghouse の動向」日本評価学会『日本評価研究』第6巻第1号、pp.31-41
- 土屋恵司 2006「2001年初等中等教育改正法（NCLB 法）の施行状況と問題点」『外国の立法』227 pp.129-135
- 松尾知明 2010『アメリカの現代教育改革ースタンダードとアカウンタビリティの光と影ー』東信堂
- 矢野裕俊 2003「アメリカにおける学力問題ー基準の設定とアカウンタビリティがもたらすものー」比較教育学研究 第29号
- Boruch, R. and R.Herman 2007 "Chapter 3. What Works Clearinghouse, United States" in OECD "Knowledge Management- Evidence in Education- Linking Research and Policy", pp.53-61=（「第3章 WWC 情報センター（米国）」『教育とエビデンス』明石書店、2009, pp.91-104）
- Donaldson, S.I and Christie, C.A. 2004 "The 2004 Claremont Debate: Lipsey vs. Scriven: Determining Causality in Program Evaluation & Applied Research: Should Experimental Evidence Be the Gold Standard?" *The Journal of Multidisciplinary Evaluation* (No.3 Oct.05) . The Evaluation Center, Western Michigan University.
- Herman, R., Boruch, R., Powerll, R., Fleischman, S., and Maynard R. 2006 "Overcoming the Challenges: A Response to Alan Shoenfeld's "What Doesn't Work." *Educational Researcher* 35(2) 22-23
- Kilpatrick, J. 2001 "Understanding Mathematical Literacy: The Contribution of Research" *Educational Studies in Mathematics* 47: 101-116
- Pope, C., N.Mays, and J.Popay. 2007 "Synthesizing Qualitative and Quantitative Health Evidence: A Guide to Methods", Open University Press =（伊藤景一・北素子（監訳）2009『質的研究と量的研究のエビデンスの統合：ヘルスケアにおける研究・実践・政策への活用』医学書院
- Slavin, R.E. 2008 "Perspectives on Evidence-Based Research in Education: What Works? Issues in Synthesizing Educational Program Evaluations" *Educational Researcher*, Vol.37, No.1, pp.5-14
- Shoenfeld, A. 2004 "The Math Wars" *Educational Policy* 18(1) 253-286
- Shoenfeld, A. 2006 "What Doesn't Work: The Challenge and Failure of the What Works Clearinghouse to Conduct Meaningful Reviews of Studies of Mathematics Curricula" *Educational Researcher* 35(2) 13-21
- U.S. Department of Education, Strategic Plan 2002-2007, Washington, D.C. , 2002
- U.S. Department of Education, Strategic Plan 2007-2012, Washington, D.C. , 2007